Statistical Analysis of Data from Sensitive Question Techniques

Ben Jann

University of Bern, jann@soz.unibe.ch

Minikonferenz des quantitativen Methodenzentrums der Universität Leipzig zum Thema „Asking Sensitive Questions: Theory and Data Collection Methods“

7.-9. Juni 2012

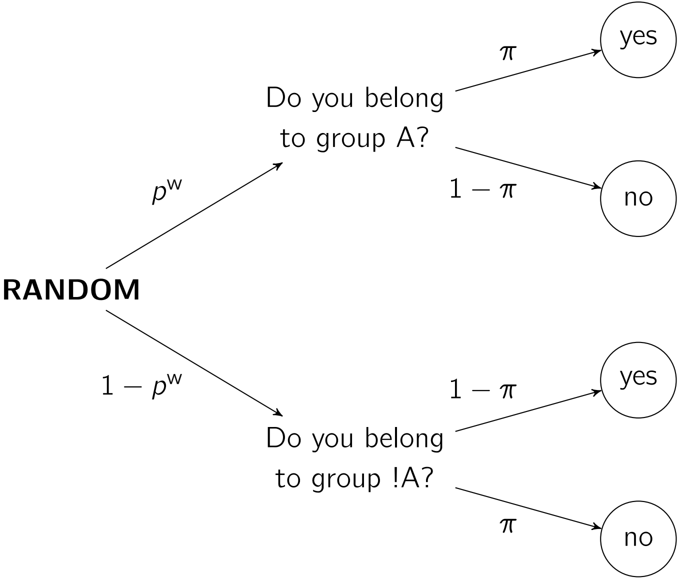
Outline

- Introduction
- Prevalence estimators for various RRT schemes
- Generalized regression estimators for RRT
- Cheating correction in RRT
- Analysis of Item Count Data

Introduction

- Various special techniques have been proposed to collect data for sensitive questions.
- The basic idea of these techniques is to anonymize answers by adding noise to the data (e.g. replacing some of the answers by random answers, aggregating answers from several questions)
- As long as the properties of the misclassification procedure are known, the statistical distribution of the sensitive question can be recovered.
- Some of these techniques are
  - Randomized Responde Technique (RRT) in various variants
    - Warner, unrelated question, forced-response, Mangat, Kuk, Crosswise Model, ...
  - Item Count Technique (ICT) a.k.a. List Experiment

Warner's RRT (Warner 1965)



Warner's RRT (Warner 1965)

- "Group A" is the sensitive group, i.e. belongig to group A is equivalent to answering "yes" to the sensitive question (SQ = 1).
- Point estimate for $\pi = \Pr(\text{"belongs to group A"}) = \Pr(SQ = 1)$?

$$\Pr(\text{"yes"}) = \lambda = p^{\mathrm{w}}\pi + (1 - p^{\mathrm{w}})(1 - \pi)$$

$$\pi = \frac{\lambda + p^{\mathrm{w}} - 1}{2p^{\mathrm{w}} - 1}, \quad p^{\mathrm{w}} \neq 0.5$$

$$\hat{\lambda} = \frac{1}{n}\sum_{i=1}^{n} y_i \quad \text{where} \quad y_i = \begin{cases} 1 & \text{if "yes"} \\ 0 & \text{if "no"} \end{cases}$$

$$\hat{\pi} = \frac{\hat{\lambda} + p^{\mathrm{w}} - 1}{2p^{\mathrm{w}} - 1}$$

Warner's RRT (Warner 1965)

Sampling variance of $\hat{\pi}$?

Delta method:

$$\text{Var}\{f(x)\} = \left(\frac{df(x)}{dx}\right)^2 \text{Var}(x)$$

if $f(x)$ is a linear transformation.

$$f(\hat{\lambda}) = \frac{\hat{\lambda} + p^{\text{w}} - 1}{2p^{\text{w}} - 1} \quad \Rightarrow \quad f' = \frac{1}{2p^{\text{w}} - 1}$$

$$\widehat{\text{Var}}(\hat{\lambda}) = \frac{\hat{\lambda}(1 - \hat{\lambda})}{n}$$

$$\widehat{\text{Var}}(\hat{\pi}) = \frac{\hat{\lambda}(1 - \hat{\lambda})}{n(2p^{\text{w}} - 1)^2} = \frac{\hat{\pi}(1 - \hat{\pi})}{n} + \frac{p^{\text{w}}(1 - p^{\text{w}})}{n(2p^{\text{w}} - 1)^2}$$

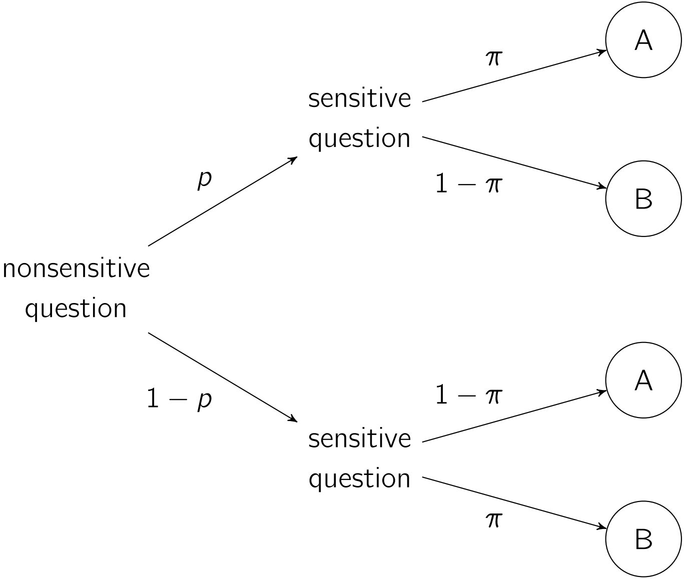
Crosswise Model (Yu et al. 2008)

- Ask a sensitive question and a nonsensitive question and let the respondent indicate whether ...
  - **A** the answers to the questions are the same (both "yes" or both "no")
  - **B** the answers are different (one "yes", the other "no")

		nonsensitive question: no	nonsensitive question: yes
sensitive question	no	**A**	**B**
	yes	**B**	**A**

- Assumtion: The two questions are uncorrelated.
- $p = \Pr(\text{"yes"})$ of the nonsensitive question must not be 0.5.

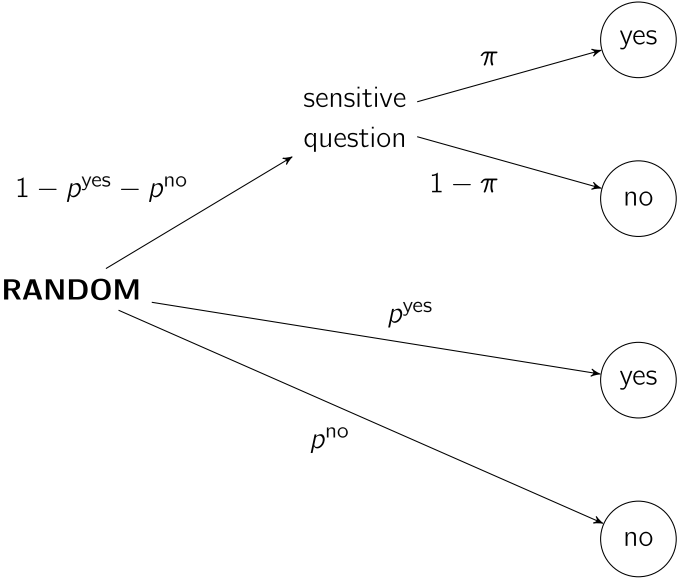
Crosswise Model (Yu et al. 2008)



Crosswise Model (Yu et al. 2008)

- The Crosswise Model is formally equivalent to Warner's RRT with $p^{\mathrm{w}} = p$.

Forced Response RRT (Boruch 1971)



Forced Response RRT (Boruch 1971)

$$\Pr(\text{“yes”}) = \lambda = (1 - p^{\text{yes}} - p^{\text{no}})\pi + p^{\text{yes}}$$

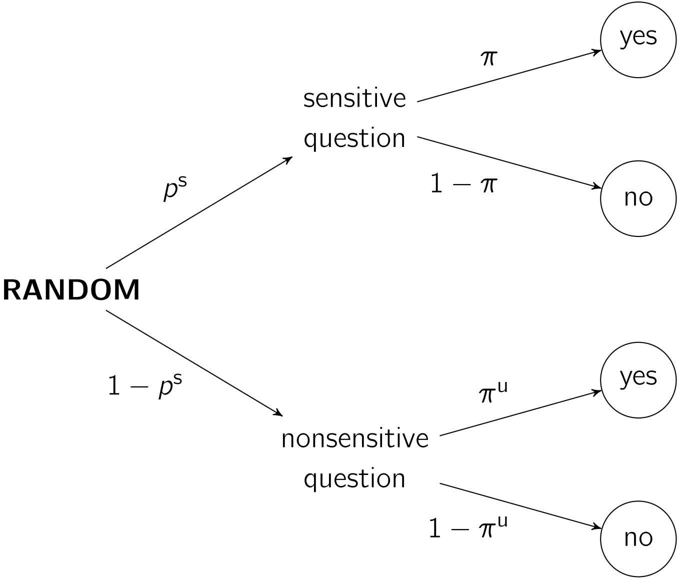
Hence

$$\hat{\pi} = \frac{\hat{\lambda} - p^{\text{yes}}}{1 - p^{\text{yes}} - p^{\text{no}}}$$

and

$$\widehat{\text{Var}}(\hat{\pi}) = \frac{\hat{\lambda}(1 - \hat{\lambda})}{n(1 - p^{\text{yes}} - p^{\text{no}})^2}$$

Unrelated Question RRT (Horvitz et al. 1967)



Unrelated Question RRT (Horvitz et al. 1967)

$$\Pr(\text{"yes"}) = \lambda = p^{\mathrm{s}}\pi + (1 - p^{\mathrm{s}})\pi^{\mathrm{u}}$$

Let

$$p^{\mathrm{s}} = 1 - p^{\mathrm{yes}} - p^{\mathrm{no}}, \quad \pi^{\mathrm{u}} = \frac{p^{\mathrm{yes}}}{p^{\mathrm{yes}} + p^{\mathrm{no}}}$$

then

$$\begin{aligned}\lambda &= (1 - p^{\mathrm{yes}} - p^{\mathrm{no}})\pi + (1 - (1 - p^{\mathrm{yes}} - p^{\mathrm{no}}))\frac{p^{\mathrm{yes}}}{p^{\mathrm{yes}} + p^{\mathrm{no}}} \\ &= (1 - p^{\mathrm{yes}} - p^{\mathrm{no}})\pi + p^{\mathrm{yes}}\end{aligned}$$

Unrelated Question RRT (Horvitz et al. 1967)

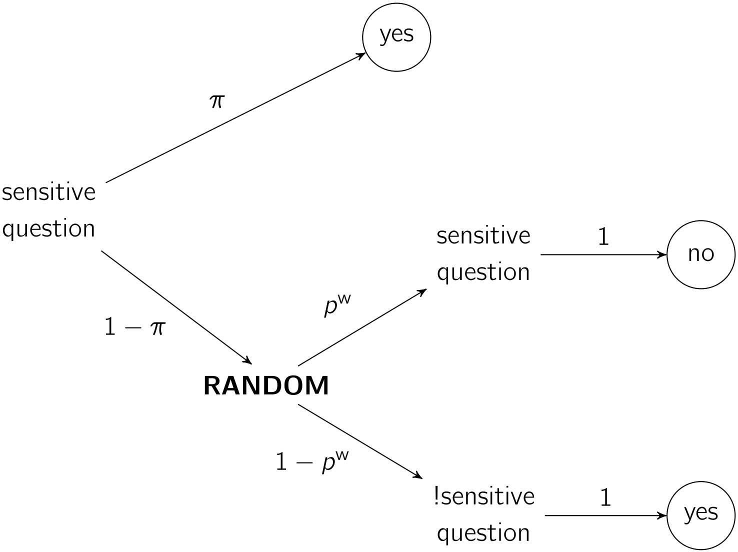
- Hence, if π^{u} is known, the Unrelated Question RRT is formally equivalent the Forced Response RRT with

$$p^{\mathrm{yes}} = (1 - p^{\mathrm{s}})\pi^{\mathrm{u}}, \quad p^{\mathrm{no}} = (1 - p^{\mathrm{s}})(1 - \pi^{\mathrm{u}})$$

- If π^{u} is unkown, it has to be estimated from a control sample. This does not change the formula for the point estimate, but it has consequences for the sampling variance (increase). Use bootstrap for variance estimation in this case.
- Alternatively, here's the variance formula (assuming that π^{u} is estimated using an independent sample):

$$\hat{\pi} = \frac{1}{p^{\mathrm{s}}}\hat{\lambda} - \frac{1 - p^{\mathrm{s}}}{p^{\mathrm{s}}}\hat{\pi}^{\mathrm{u}} \Rightarrow \widehat{\mathrm{Var}}(\hat{\pi}) = \left(\frac{1}{p^{\mathrm{s}}}\right)^2 \widehat{\mathrm{Var}}(\hat{\lambda}) + \left(\frac{1 - p^{\mathrm{s}}}{p^{\mathrm{s}}}\right)^2 \widehat{\mathrm{Var}}(\hat{\pi}^{\mathrm{u}})$$

Mangat's RRT (Mangat 1994)



Mangat's RRT (Mangat 1994)

$$\Pr(\text{“red”}) = \lambda = \pi p_1 + (1-\pi)p_2$$

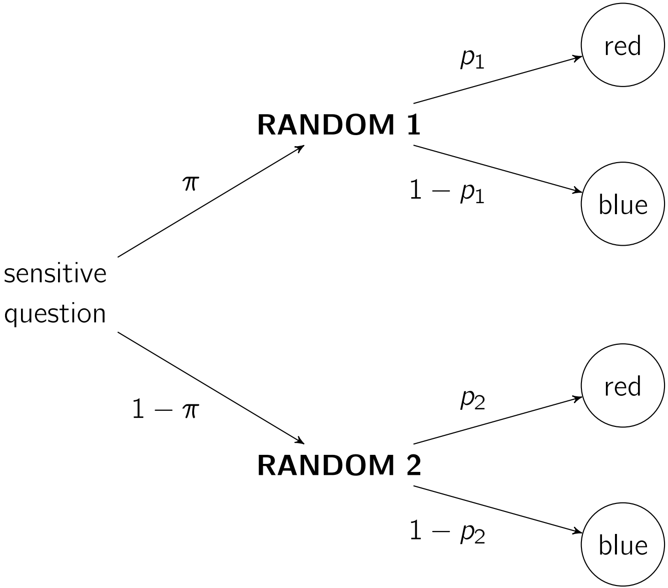
Hence

$$\hat{\pi} = \frac{\hat{\lambda} - p_2}{p_1 - p_2}, \quad p_1 \neq p_2$$

and

$$\widehat{\text{Var}}(\hat{\pi}) = \frac{\hat{\lambda}(1-\hat{\lambda})}{n(p_1 - p_2)^2}$$

Kuk's RRT (Kuk 1990)



Kuk's RRT (Kuk 1990)

$$\Pr(\text{“yes”}) = \lambda = \pi + (1-\pi)(1-p^{\mathrm{w}})$$

Hence

$$\hat{\pi} = \frac{\hat{\lambda} + p^{\mathrm{w}} - 1}{p^{\mathrm{w}}}$$

and

$$\widehat{\mathrm{Var}}(\hat{\pi}) = \frac{\hat{\lambda}(1-\hat{\lambda})}{n(p^{\mathrm{w}})^2} = \frac{\hat{\pi}(1-\hat{\pi})}{n} + \frac{(1-\hat{\pi})(1-p^{\mathrm{w}})}{np^{\mathrm{w}}}$$

Generalized regression estimator for RRT

- Let

Y_i	response ($Y_i = 1$ if "yes" in RRT or "A" in CM, else $Y_i = 0$)
λ_i	probability of $Y_i = 1$
π_i	(unknown) prevalence of sensitive item
p_i^{w}	probability of the non-negated question in Warner's RRT (prevalence of nonsensitive item in CM)
p_i^{yes}	probability of a forced "yes"
p_i^{no}	probability of a forced "no"

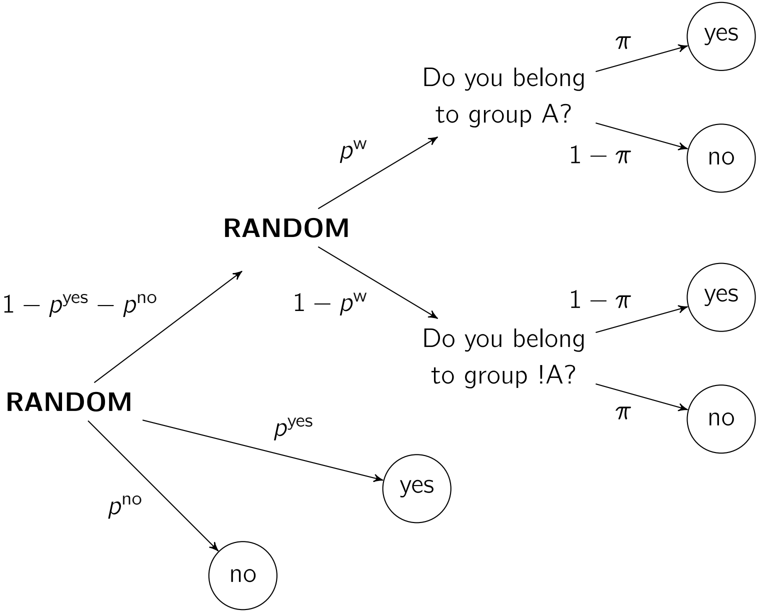
- Then

$$\lambda_i = (1 - p_i^{\mathrm{yes}} - p_i^{\mathrm{no}}) p_i^{\mathrm{w}} \pi_i + (1 - p_i^{\mathrm{yes}} - p_i^{\mathrm{no}})(1 - p_i^{\mathrm{w}})(1 - \pi_i) + p_i^{\mathrm{yes}}$$

and hence

$$\pi_i = \frac{\lambda_i - (1 - p_i^{\mathrm{yes}} - p_i^{\mathrm{no}})(1 - p_i^{\mathrm{w}}) - p_i^{\mathrm{yes}}}{(2p_i^{\mathrm{w}} - 1)(1 - p_i^{\mathrm{yes}} - p_i^{\mathrm{no}})}$$

Generalized regression estimator for RRT



Generalized regression estimator for RRT

- By parametrizing π_i we can formulate regression models.
- For example, assuming $\pi_i = X_i'\beta$, we can estimate β by applying least squares regression to a transformed response variable

$$\tilde{Y}_i = \frac{Y_i - (1 - p_i^{\text{yes}} - p_i^{\text{no}})(1 - p_i^{\text{w}}) - p_i^{\text{yes}}}{(2p_i^{\text{w}} - 1)(1 - p_i^{\text{yes}} - p_i^{\text{no}})}$$

- This is because

$$E(\text{SQ} = 1|X_i) = \frac{E(Y_i|X_i) - (1 - p_i^{\text{yes}} - p_i^{\text{no}})(1 - p_i^{\text{w}}) - p_i^{\text{yes}}}{(2p_i^{\text{w}} - 1)(1 - p_i^{\text{yes}} - p_i^{\text{no}})}$$

Generalized regression estimator for RRT

- More resonable might be to assume a functional form such as $\ln(\pi_i/(1-\pi_i)) = X_i'\beta$ (logit), i.e. $\pi_i = e^{X_i'\beta}/(1+e^{X_i'\beta})$.
- In this case, we can derive the log likelihood as

$$
\begin{aligned}
\ln L &= \sum_{i=1}^{n} \left[Y_i \ln(\lambda_i) + (1-Y_i)\ln(1-\lambda_i) \right] \\
&= \sum_{i=1}^{n} \left[Y_i \ln(R_i) + (1-Y_i)\ln(S_i) - \ln(1+e^{X_i\beta}) \right]
\end{aligned}
$$

with

$$
\begin{aligned}
R_i &= c_i + q_i e^{X_i'\beta} & c_i &= (1-p_i^{\text{yes}}-p_i^{\text{no}})(1-p_i^{\text{w}}) + p_i^{\text{yes}} \\
S_i &= (1-c_i) + (1-q_i)e^{X_i'\beta} & q_i &= (1-p_i^{\text{yes}}-p_i^{\text{no}})p_i^{\text{w}} + p_i^{\text{yes}}
\end{aligned}
$$

and estimate β using maximum likelihood methods.

Two Stata commands

- Least-squares estimation with $\pi_i = X_i'\beta$ (Jann 2008):

  `rrreg` *depvar* [*indepvars*] [*if*] [*in*] [*weight*] [, *regress_options*
  `pwarner(`# | *varname*`)` `pyes(`# | *varname*`)` `pno(`# | *varname*`)`]

- Maximum likelihood estimation with $\pi_i = e^{X_i'\beta}/(1 + e^{X_i'\beta})$ (Jann 2005):

  `rrlogit` *depvar* [*indepvars*] [*if*] [*in*] [*weight*] [, *logit_options*
  `pwarner(`# | *varname*`)` `pyes(`# | *varname*`)` `pno(`# | *varname*`)`]

- `rrlogit` may make more sense in terms of functional form. However, `rrreg` is more robust, especially if there is noncompliance with the RRT procedure.

Example

```
. use gr/rrt07
(Sensitive Questions Online Survey 2007)

. fre grp

grp — Experimental group
```

		Freq.	Percent	Valid	Cum.
Valid	1 direct	609	38.42	38.42	38.42
	2 manual coin toss	169	10.66	10.66	49.09
	3 electronic coin toss	188	11.86	11.86	60.95
	4 banknote with phone	227	14.32	14.32	75.27
	5 banknote without phone	190	11.99	11.99	87.26
	6 phone number	202	12.74	12.74	100.00
	Total	1585	100.00	100.00	

```
. generate rrt = inrange(grp,2,6)

. regress keepchange if rrt==0
```

Source	SS	df	MS
Model	0	0	.
Residual	149.748355	607	.246702397
Total	149.748355	607	.246702397

Number of obs =	608
F(0, 607) =	0.00
Prob > F =	.
R-squared =	0.0000
Adj R-squared =	0.0000
Root MSE =	.49669

keepchange	Coef.	Std. Err.	t	P>|t|	[95% Conf. Interval]	
_cons	.5608553	.0201435	27.84	0.000	.5212959	.6004147

(Data fromm Coutts and Jann 2011)

Example

```
. rrreg keepchange if rrt==1, pyes(0.5)

Randomized response regression                    Number of obs  =       927
                                                  F(   0,    926) =      0.00
                                                  Prob > F        =         .
                                                  R-squared       =    0.0000
                                                  Adj R-squared   =    0.0000
                                                  Root MSE        =    0.8046

------------------------------------------------------------------------------
  keepchange |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       _cons |   .5943905   .0264269    22.49   0.000      .542527    .646254
------------------------------------------------------------------------------

Pr(non-negated question) = 1
Pr(surrogate "yes")      = 0.5
Pr(surrogate "no")       = 0
```

Example

```
. generate pyes = cond(rrt==1, 0.5, 0)
. rrreg keepchange rrt, pyes(pyes)
Randomized response regression                     Number of obs   =       1535
                                                   F(   1,   1533) =       0.84
                                                   Prob > F        =     0.3581
                                                   R-squared       =     0.0006
                                                   Adj R-squared   =    -0.0001
                                                   Root MSE        =     0.6991

------------------------------------------------------------------------------
  keepchange |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         rrt |   .0335352   .0364839     0.92   0.358    -.0380285    .1050989
       _cons |   .5608553   .0283522    19.78   0.000      .505242    .6164685
------------------------------------------------------------------------------
Pr(non-negated question) = 1
Pr(surrogate "yes")      = pyes
Pr(surrogate "no")       = 0
```

Example

```
. rrreg keepchange rrt highschool, pyes(pyes)

Randomized response regression                     Number of obs   =       1535
                                                   F(   2,    1532) =       4.53
                                                   Prob > F        =     0.0110
                                                   R-squared       =     0.0059
                                                   Adj R-squared   =     0.0046
                                                   Root MSE        =     0.6975

------------------------------------------------------------------------------
  keepchange |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         rrt |   .0347934   .0364012     0.96   0.339     -.036608    .1061948
  highschool |   .1055695   .0368606     2.86   0.004     .0332669    .1778721
       _cons |   .4936589   .0367501    13.43   0.000     .4215731    .5657446
------------------------------------------------------------------------------

Pr(non-negated question) = 1
Pr(surrogate "yes")      = pyes
Pr(surrogate "no")       = 0
```

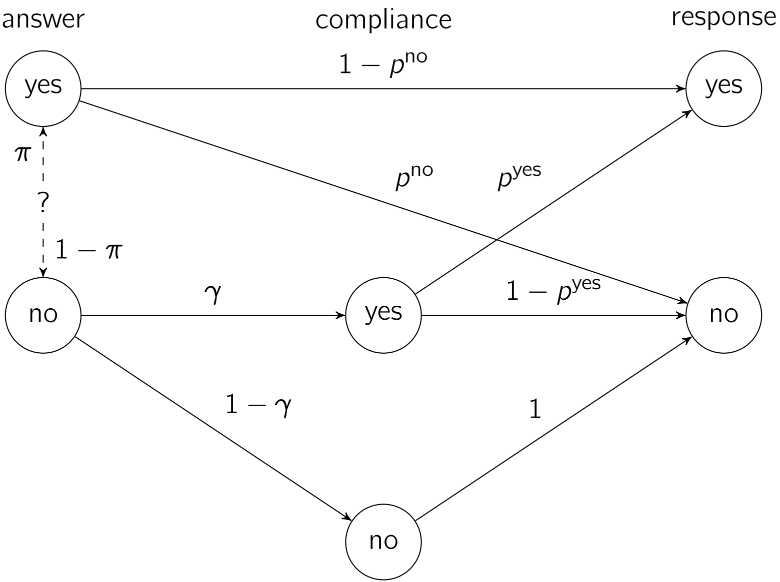
Example

```
. generate rrtXhs = rrt*highschool

. rrreg keepchange rrt highschool rrtXhs, pyes(pyes)

Randomized response regression                    Number of obs   =       1535
                                                  F(   3,   1531) =       3.32
                                                  Prob > F        =     0.0193
                                                  R-squared       =     0.0065
                                                  Adj R-squared   =     0.0045
                                                  Root MSE        =     0.6975

------------------------------------------------------------------------------
  keepchange |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         rrt |   .0799137   .0599938     1.33   0.183    -.0377651    .1975924
  highschool |   .1489237    .058808     2.53   0.011      .033571    .2642765
      rrtXhs |   -.071412   .0754755    -0.95   0.344    -.2194583    .0766344
       _cons |   .4660633   .0469181     9.93   0.000     .3740329    .5580938
------------------------------------------------------------------------------

Pr(non-negated question) = 1
Pr(surrogate "yes")      = pyes
Pr(surrogate "no")       = 0
```

A little bit of magic: Cheating correction in RRT

- In many RRT designs, the "self-protective no" bias can occur.
- In these designs, some of the respondents are instructed to answer "yes" by the randomization device, even though the sensitive item does not apply to them.
- There is evidence that these respondents often deviate from the instructions and answer "no".
- Such non-compliance introduces a large bias to RRT estimates. It is noteworthy that this bias does not come from respondents who did commit the sensitive behavior and want to conceal it. It comes from respondents who did not and don't want it to look like they did.
- In a standard design, it is not possible to account for such "cheaters". However, if the RRT design parameters are variied, this variation can be used to identify the proportion of cheaters and correct the estimates.

A little bit of magic: Cheating correction in RRT



A little bit of magic: Cheating correction in RRT

- Assumptions:
  - There is random variation in p^{yes} and p^{no} between respondents.
  - π and γ do not depend on p^{yes} and p^{no} (which may be justified if the variation in p is small)
  - Respondents do not say "yes" if instructed to say "no" by the randomization device.
- π and γ can then be estimated using the following log likelihood:

$$\ln L = \sum_{i=1}^{n} Y_i \ln(\ell_i) + (1 - Y_i) \ln(1 - \ell_i)$$

with

$$\ell_i = \pi_i(1 - p_i^{\text{no}} - \gamma p_i^{\text{yes}}) + \gamma p_i^{\text{yes}}$$

A little bit of magic: Analysis

```
program define rrcheat_lf
    args lnf theta1 cheat
    local p1 $rrcheat_pyes
    local p2 $rrcheat_pno
    quietly replace `lnf´ = cond($ML_y1, ///
        ln(`theta1´ * (1 - `p2´ - (1-`cheat´)*`p1´) + (1-`cheat´)*`p1´), ///
        ln(1 - (`theta1´ * (1 - `p2´ - (1-`cheat´)*`p1´) + (1-`cheat´)*`p1´)))
end
forv i = 1/5 {
    local depvar: word `i´ of $sqvar
    global rrcheat_pyes pyesQ`i´
    global rrcheat_pno  pnoQ`i´
    ml model lf rrcheat_lf (`depvar´: `depvar´ = ) /cheat if RRT==1
    ml maximize
    eststo `depvar´
}
esttab, nonumb  nostar mti se b(1) transform(100*@ 100) ///
    eqlab(none) coef(main:_cons "RRT adjusted" cheat:_cons "Cheaters")
```

A little bit of magic: Results

	copy	notes	drugs	partial	severe
RRT adjusted	17.9	12.0	16.7	14.3	6.7
	(6.5)	(6.1)	(5.6)	(6.6)	(5.9)
Cheaters	-9.5	-3.6	88.9	54.3	36.1
	(36.1)	(31.9)	(36.9)	(40.1)	(31.8)
N	2855	2855	2849	2105	2104

Standard errors in parentheses

(Data from Höglinger et al. 2012)

Unadjusted results for comparison:

	copy	notes	drugs	partial	severe
DQ	17.5	8.8	3.4	2.5	1.5
	(1.2)	(0.9)	(0.6)	(0.6)	(0.5)
RRT	19.6	12.7	0.6	4.2	-0.6
	(1.2)	(1.1)	(1.0)	(1.2)	(1.1)
CM	27.2	15.0	9.9	8.2	3.0
	(2.0)	(1.9)	(1.9)	(2.1)	(2.0)

Analysis of Item Count Data

- Item Count Design (Droitcour et al. 1991):

group A (short list)	group B (long list)
nonsensitive item 1	nonsensitive item 1
nonsensitive item 2	nonsensitive item 2
nonsensitive item 3	nonsensitive item 3
	sensitive item

- How many items do apply to you?
- Two randomized groups, one with the short list, one with the long list (single list design).

Analysis of Item Count Data

- Estimate of the probability of the sensitive item $\pi = \Pr(SQ = 1)$?
- Mean difference between the two groups:

$$\hat{\pi} = \bar{y}^{\text{LL}} - \bar{y}^{\text{SL}} = \frac{1}{n^{\text{B}}} \sum_{i \in \text{B}} y_i - \frac{1}{n^{\text{A}}} \sum_{i \in \text{A}} y_i$$

- Variance of $\hat{\pi}$?

$$\text{Var}(\hat{\pi}) = \text{Var}(\bar{y}^{\text{LL}}) + \text{Var}(\bar{y}^{\text{SL}})$$

Analysis of Item Count Data

- Double list design:
  - Both groups answer to two sets of items. In one group, the sensitive item is paired with the first set of nonsensitive items, in the other group the sensitive item is paired with the second set of nonsensitive items.

Set 1:	group A	group B
	nonsensitive item 1	nonsensitive item 1
	nonsensitive item 2	nonsensitive item 2
	nonsensitive item 3	nonsensitive item 3
		sensitive item

Set 2:	group A	group B
	nonsensitive item 4	nonsensitive item 4
	nonsensitive item 5	nonsensitive item 5
	nonsensitive item 6	nonsensitive item 6
	sensitive item	

Analysis of Item Count Data

$$\hat{\pi}_1 = \bar{y}^{\mathrm{LL1}} - \bar{y}^{\mathrm{SL1}}, \quad \hat{\pi}_2 = \bar{y}^{\mathrm{LL2}} - \bar{y}^{\mathrm{SL2}}$$

$$\begin{aligned}
\hat{\pi} &= \frac{\hat{\pi}_1 + \hat{\pi}_2}{2} = \frac{(\bar{y}^{\mathrm{LL1}} - \bar{y}^{\mathrm{SL1}}) + (\bar{y}^{\mathrm{LL2}} - \bar{y}^{\mathrm{SL2}})}{2} \\
&= \frac{(\bar{y}^{\mathrm{LL1}} - \bar{y}^{\mathrm{SL2}}) + (\bar{y}^{\mathrm{LL2}} - \bar{y}^{\mathrm{SL1}})}{2} \\
&= \frac{\frac{1}{n^{\mathrm{B}}}\sum_{i \in \mathrm{B}}(y_{1i} - y_{2i}) + \frac{1}{n^{\mathrm{A}}}\sum_{i \in \mathrm{A}}(y_{2i} - y_{1i})}{2}
\end{aligned}$$

$$\begin{aligned}
\mathrm{Var}(\hat{\pi}) &= \frac{\mathrm{Var}(\hat{\pi}_1) + \mathrm{Var}(\hat{\pi}_2) - 2\mathrm{Cov}(\hat{\pi}_1, \hat{\pi}_2)}{4} \\
&= \frac{\mathrm{Var}(\bar{y}^{\mathrm{LL1}} - \bar{y}^{\mathrm{SL2}}) + \mathrm{Var}(\bar{y}^{\mathrm{LL2}} - \bar{y}^{\mathrm{SL1}})}{4}
\end{aligned}$$

Analysis of Item Count Data

- Regression model for single list design:
  - Estimate β by applying least-squares regression (with robust standard errors) to

    $$Y_i = (LL_i \cdot X_i)'\beta + X_i'\gamma + \epsilon_i$$

    (For more sophisticated approaches see Glynn 2010, Imai 2010, Blair and Imai 2012.)
- Regression model for single list design:
  - Approach 1: estimate separate models (as above) for Y_1 and Y_2, combine estimates using `suest` to obtain joint variance matrix, compute average coefficients using `lincom`
  - Approach 2: estimate a system of equations (e.g. using `sureg`) for Y_1 and Y_2 with the contraint that the coefficients are the same

Conclusions

- Suitable methods for basic analysis of data from sensitive question techniques (SQT) are easy to derive.
- Canned software exists for various RRT designs.
- Outlook
  - Add support for Mangat's RRT, Kuk's RRT, ...
  - Canned software for Item Count Data
  - The presented methods treat the SQT-data as the dependent variable. What if SQT-variables are used as predictors?
  - Correlations among SQT-variables? Analysis of multiple SQT-items?
  - More sophisticated designs/methods for cheating correction?

References I

- Blair, G., K. Imai. 2012. Statistical Analysis of List Experiments. *Political Analysis* 20(1): 47–77.
- Boruch, R.F. 1971. Assuring Confidentiality of Responses in Social Research: A Note on Strategies. *The American Sociologist* 6(4): 308-311.
- Coutts, E., B. Jann. 2011. Sensitive Questions in Online Surveys: Experimental Results for the Randomized Response Technique (RRT) and the Unmatched Count Technique (UCT). *Sociological Methods & Research* 40(1): 169–193.
- Droitcour, J., R.A. Caspar, M.L. Hubbard, T.L. Parsely, W. Vischer, T.M. Ezzati. 1991. The item count technique as a method of indirect questioning: A review of its development and a case study application. Pp. 185-210 in: P. Biemer, R.M. Groves, L.E. Lyberg, N.A. Mathiowetz, S. Sudman (eds.), Measurement errors in surveys. New York: Wiley.
- Glynn, A.N. 2010. What Can We Learn with Statistical Truth Serum? Design and Analysis of the List Experiment. Harvard University.

References II

- Höginger, M., B. Jann, A. Diekmann. 2012. Sensitive Questions in Online Surveys: An Experimental Comparison of the Randomized Response Technique and the Crosswise Model. ETH Zurich and University of Bern.
- Horvitz, D.G., B.V. Shah, W.R. Simmons. 1967. The unrelated question randomized response model. Pp. 65-72 in: Proceedings in the Social Science Section. Baltimore: American Statistical Association.
- Imai, K. 2011. Multivariate regression analysis for the item count technique. *Journal of the American Statistical Association* 106:407–16.
- Jann, B. 2005. rrlogit: Stata module to estimate logistic regression for randomized response data. Statistical Software Components S456203, Boston College Department of Economics, revised 12 May 2011. Available from http://ideas.repec.org/c/boc/bocode/s456203.html
- Jann, B. 2008. rrreg: Stata module to estimate linear probability model for randomized response data. Statistical Software Components S456962, Boston College Department of Economics, revised 12 May 2011. Available from http://ideas.repec.org/c/boc/bocode/s456962.html

References III

- Kuk, A.Y.C. 1990. Asking sensitive questions indirectly. *Biometrika* 77(2): 436-438.
- Mangat, N.S. 1994. An Improved Randomized Response Strategy. *Journal of the Royal Statistical Society. Series B (Methodological)* 56(1): 93–95.
- Warner, S.L. 1965. Randomized-response: A survey technique for eliminating evasive answer bias. *Journal of the American Statistical Association* 60(309):63–69.
- Yu, J.-W., G.-L. Tian, M.-L. Tang. 2008. Two new models for survey sampling with sensitive characteristic: design and analysis. *Metrika* 67(3):251–263.